Exploiting Sensorimotor Coordination for Learning to Recognize Objects [*]

Yohannes Kassahun[†], Mark Edgington[†], Jose de Gea[†] and Frank Kirchner[†] [⋆]
Robotics Group
[†] University of Bremen
[⋆] German Research Center for Artificial Intelligence (DFKI)
Robert-Hooke-Str. 5, D-28359, Bremen, Germany
{kassahun, edgimar, jdegea, frank.kirchner}@informatik.uni-bremen.de

Abstract

In this paper we present a system which learns to recognize objects through interaction by exploiting the principle of sensorimotor coordination. The system uses a learning architecture which is composed of reactive and deliberative layers. The reactive layer consists of a database of behaviors that are modulated to produce a desired behavior. In this work we have implemented and installed in our architecture an object manipulation behavior inspired by the concept that infants learn about their environment through manipulation. While manipulating objects, both proprioceptive data and exteroceptive data are recorded. Both of these types of data are combined and statistically analyzed in order to extract important parameters that distinctively describe the object being manipulated. This data is then clustered using the standard k-means algorithm and the resulting clusters are labeled. The labeling is used to train a radial basis function network for classifying the clusters. The performance of the system has been tested on a kinematically complex walking robot capable of manipulating objects with two legs used as arms, and it has been found that the trained neural network is able to classify objects even when only partial sensory data is available to the system. Our preliminary results demonstrate that this method can be effectively used in a robotic system which learns from experience about its environment.

1 Introduction

Recently, the psychological point of view that grants the body a more significant role in cognition has also gained attention in spatial cognition theory. Proponents of this approach claim that we have to deal with and understand a body that needs a mind to make it function instead of a mind that works on abstract problems [Wilson, 2002]. These ideas differ quite radically from the traditional approach that describes a cognitive process as an abstract information processing task in which the real physical connections to the outside world are of only sub-critical importance, and are sometimes discarded as mere informational encapsulated plug-ins [Fodor, 1983]. Most theories in cognitive psychology take this traditional approach and have tried to describe the process of human thinking in terms of propositional knowledge. At the same time, artificial intelligence research has been dominated by methods of abstract symbolic processing, even when researchers have used robotic systems to implement them [Nilsson, 1984]. Ignoring sensorimotor influences on cognitive ability is in sharp contrast to the research of William James [James, 1890] and others (see [Prinz, 1987] for a review) that describe theories of cognition based on motor acts, or a theory of cognitive function emerging from seminal research on sensorimotor abilities by Jean Piaget [Wilson, 2002] and the theory of affordances by [Gibson, 1977]. In the 1980s the linguist Lakoff and the philosopher Johnson [Lakoff and Johnson, 1980] put forward the idea of abstract concepts based on metaphors for bodily, physical concepts; around the same time, Brooks [Brooks, 1986] made a major impact on artificial intelligence research by his concepts of behavior based robotics, and interaction with the environment without internal representation. This concept provides an alternative to the traditional sense-reason-act cycle, and has gained wide attention ever since. As promising as these ideas seem to be at first glance, one has to carefully evaluate what exact claims can be made and how these can be evaluated. Wilson identifies six viewpoints for the new so-called embodied cognition approach [Wilson, 2002]:

1. Cognition is situated: All Cognitive activity takes part in the context of a real world environment.
2. Cognition is time pressured: How does cognition work under the pressures of real time interaction with the environment
3. *Off-loading of cognitive work to the environment: Limits of our information processing capabilities demand for off-loading.*
4. The environment is part of the cognitive system: because of dense and continuous information flow between the mind and the environment it is not meaningful to study just the mind.
5. *Cognition is for action: Cognitive mechanisms (perception/memory, etc.) must be understood in their ultimate*

[*] This work is supported by the SFB/TR 8 DFG project, which is duly acknowledged.

contribution to situation appropriate behavior.

6. Off-line Cognition is body-based: Even when uncoupled from the environment, the activity of the mind is grounded in mechanisms that evolved for interaction with the environment.

We have cited all six viewpoints here, as they represent an interesting perspective on the state of the art in embodied cognition. In this work we focus our attention on viewpoints 3 and 5, and we use them as a theoretical starting point for our work. To experiment with embodied cognition, we propose the use of a multifunctional four legged robot kinematically capable of walking and climbing on four legs as well as of grasping and manipulating objects with two legs used as arms. The role of manipulation acts in understanding spatial geometries and objects goes back to the idea that cognitive systems offload as much of the computational burden as possible onto the environment to understand spatial structures. Instead of generating and transforming complex mathematical models of 3-D geometries, cognitive systems use motor acts to generate multi-modal perceptual inputs, which they use to test hypotheses about the nature of the geometric structure at hand.

A prerequisite for developing higher levels of cognition is the process of sensorimotor coordination, in which the body of the system plays a central role in learning to recognize objects [Nolfi and Parisi, 1999; Pfeifer and Scheier, 1999]. Many researchers [Edelman, 1987; Beer, 1996; Nolfi, 1996; Takác, 2006] have shown that sensorimotor coordination can be exploited in solving categorization problems.

One shortcoming of most existing methods is that they are able to recognize only a limited number of objects. Additionally, most existing methods are difficult to extend. A typical application of such methods is to recognize hypothetical objects, and they are tested only in simulation or on simple robotic platforms. Their ability to scale when used on complex robots is neither known nor proven. From our viewpoint, the reasons for the shortcomings of the existing methods are twofold: First, there is no presently firm theoretical framework for studying correlations within and between sensorimotor modalities for object recognition tasks. Very few approaches apply statistical and information theoretic analyses to study the sensorimotor coordination of data taken from real robots [Lungarella *et al.*, 2005]. Second, kinematically complex robots capable of increasing the role of the body in the process of learning and recognition are not commonly used. Most of the time wheeled robots with few degrees of freedom or simulated robotic arms are used as test beds.

After the introduction of Brooks' behavior based robotics and interaction with environment [Brooks, 1986], there appears to be a growing sense of commitment to the idea that cognitive ability in a system, be it natural or artificial, has to be studied in the context of its relation to a 'kinematically competent' physical body. Therefore, in the last years we focused our research on complex legged robots which possess a rich repertoire of sensor and motor abilities [Hilljegerdes *et al.*, 2005].

In this paper we present an extensible embodied object recognition system that can be used in complex real robots that learn through interaction with the environment. The system can be easily extended to use new object-features which distinctively describe the relevant characteristics of an object to be recognized.

A work that is closely related to ours involves substrate classification on the basis of proprioceptive data [Spenneberg and Kirchner, 2005]. In this work, a legged robot named SCORPION [Kirchner *et al.*, 2002] interacts with various substrates to generate certain substrate specific sensory feedback. The results of this experiment show that this method of classifying based on proprioceptive data has a promising potential for use in terrain classification of unstructured environments. One of the important benefits of terrain classification is that it allows a terrain's traversability to be assessed given a specific robot body.

This paper is organized as follows: first, we give a short overview of the learning architecture which we have used to implement object recognition through manipulation. We then explain the manipulation behavior and the recognition method used. Next, we describe our experimental scenario and the results obtained. Finally, we provide some conclusions and a future outlook.

2 Learning Architecture

The architecture we have adopted, shown in Figure 1, is a hybrid architecture which integrates a reactive system with a higher-level deliberative system. It is suitable for controlling and integrating spatial learning and representation techniques in mobile robots, allowing them to explore and navigate in unknown environments.

2.1 Reactive System

The reactive system includes the `INPUT`, `TRANSFER`, `ACTIVATION`, and `MOTOR` modules of Figure 1. Proprioceptive and exteroceptive data produced through interaction with the environment are processed in the `INPUT` stage. This stage consists of three subsystems (see Figure 2) which work together in order to learn, distinguish and identify the perceptual states the system is in. In the manipulation based object recognition experiment, an unsupervised classifier system is implemented that considers both proprioceptive and exteroceptive data, represented by the vectors $\vec{S}_P$ and $\vec{S}_E$, respectively. This classifier identifies and labels clusters of similar sensorimotor stimuli together. Additionally, it generates a cluster probability $\vec{P}_{C,sm}(\vec{S}_P, \vec{S}_E)$. Each element of this vector represents a probability estimate of the likelihood that a set of sensorimotor inputs belongs to a labeled cluster. The cluster probability $\vec{P}_{C,sm}$ is then used to train the two remaining classifiers which classify based only on the exteroceptive sensory data vector $\vec{S}_E$ or on the proprioceptive data vector $\vec{S}_D$. These classifiers also generate $\vec{P}_C$ estimates, $\vec{P}_{C,ext}$ and $\vec{P}_{C,prop}$, which are are combined with $\vec{P}_{C,sm}$ to generate an overall cluster probability $\vec{P}_{C,input}$.

This overall cluster probability is mapped in the `TRANSFER` module to a set of motor-program activation levels. The activation levels serve to pre-activate a selected group of motor-programs. Pre-activation consists of setting

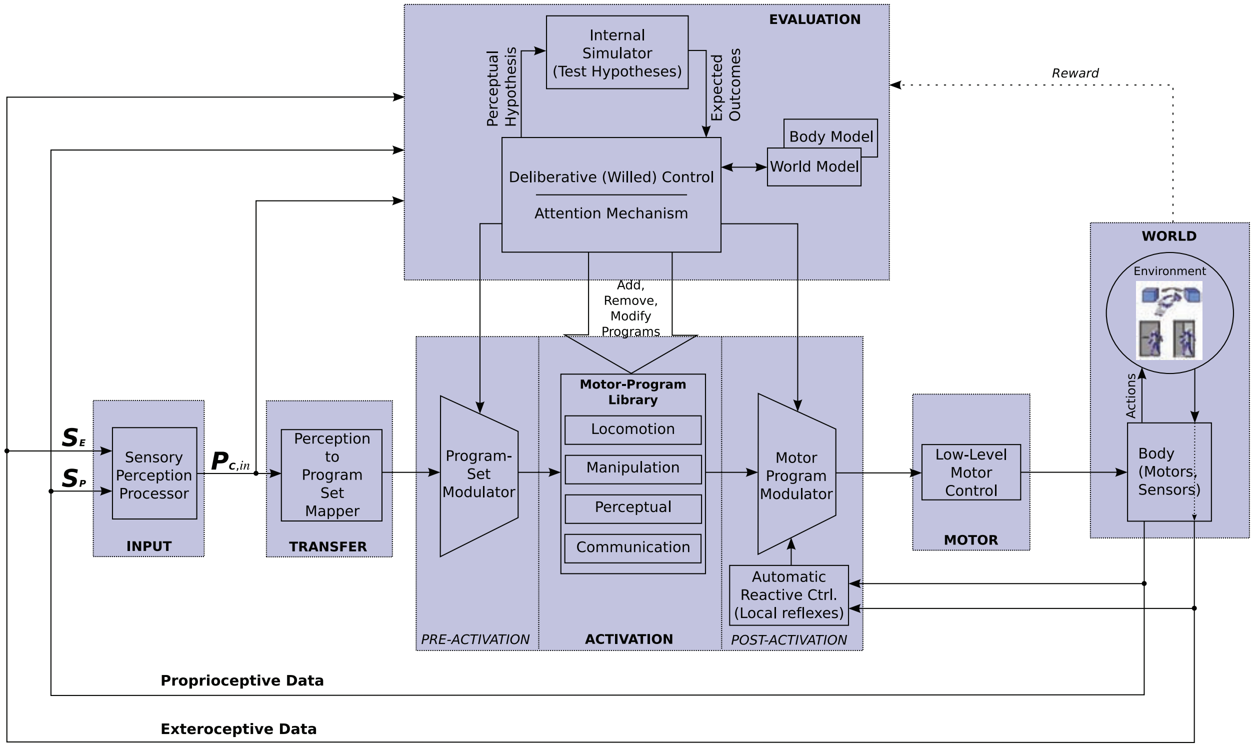

Figure 1: Overall learning architecture

all the parameters (i.e. phase-shift, amplitude, frequency, etc.) of a motor-program, as well as a weight value which is used in a further stage to merge multiple motor-programs into a resulting behavior [Kirchner *et al.*, 2002]. During the learning phase of the experiment presented here, a manipulation motor-program is made active.

The ACTIVATION stage generates a set of joint angles for each motor-program. At this point, it is possible for the output of the motor-programs to be modulated by other systems. An automatic reactive control system receives proprioceptive sensory feedback and modulates the motor-program output in order to realize local-reflex behaviors. Additionally, a higher-level deliberative system is able to influence a resulting action by modulating the motor-program output. Finally, the output from the ACTIVATION stage is fed into the MOTOR stage, which implements a believed behavior, attempting to directly control the robot's actuators. We have used the term believed behavior to indicate that there is uncertainty as to whether an intended behavior has indeed been carried out. Likelihood estimation of a successful behavior execution is made possible by comparing real perceptions with expected perceptions which come from hypotheses about how the robot's perceptions should change when a set of motor-programs has successfully been executed.

2.2 Deliberative System

The deliberative system is responsible for high-level processing of cluster probabilities, and it is in this system that the world model and body model are generated through learning. Positive and negative rewards combined with believed behavior information are used to learn the world model and body model of the robot itself. The deliberative system is also responsible for optimizing the existing motor-programs or adding new motor-programs whose resulting behavior cannot be obtained by combining the existing motor programs. Additionally, the deliberative system pre-modulates the output of the TRANSFER module to affect which motor-programs will be active, and post-modulates the output of the ACTIVATION module, modifying the properties of the motor-programs that are currently active.

3 The Recognition System

The embodied recognition system functions by manipulating objects in order to determine their specific characteristics. A manipulation motor-program has been implemented, added to the ACTIVATION module, and made active for the experiment. This motor-program uses a potential field method [Khatib, 1985] to generate a trajectory for an end-effector to reach an object. The basic idea is to create a mathematical description of a virtual potential field acting within the workspace of the manipulator. Regions in the workspace that are to be avoided are modelled by repulsive potentials (energy peaks) and the target region/point is modelled by an attractive potential (energy valley). The sum of repulsive and attractive potentials provides a representation of the workspace topology. By following the gradient (i.e. the minimum potential field at each step), a path towards the goal is generated. One fundamental difference between this method and classi-

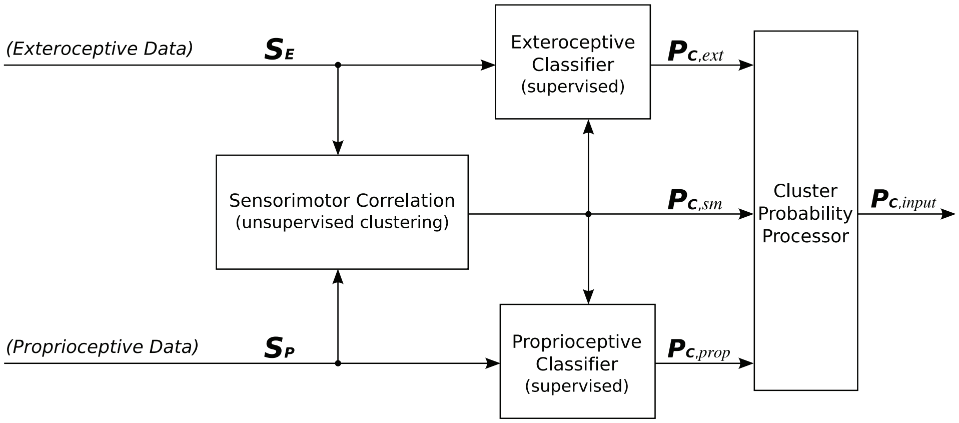
(Exteroceptive Data)
S_E
Exteroceptive Classifier (supervised)
$P_{C,ext}$
Sensorimotor Correlation (unsupervised clustering)
$P_{C,sm}$
Cluster Probability Processor
$P_{C,input}$
(Proprioceptive Data)
S_P
Proprioceptive Classifier (supervised)
$P_{C,prop}$

Figure 2: Sensory perception processor

cal path planning is that here ”planning” is not done in the usual sense. Rather, a path is incrementally computed that will end at the target position. This approach can be viewed as a reactive approach since there is no deliberation involved and it can be implemented on lower layers of control. Furthermore, this reactiveness allows us to deal with obstacles on a real-time basis, the only limitation being the time needed to detect and identify objects as obstacles or goals.

While manipulating objects, both proprioceptive data (pressure at fingertips, motor current consumption, motor angular position) and exteroceptive data (color of the object, number of corners detected on the object, number of line segments detected, and other distinctive features) are recorded. Both of these types of data are combined to form a vector $\vec{X} = [\vec{S}_P, \vec{S}_E]$. The resulting vector is statistically analyzed in order to extract important parameters that distinctively describe the object being manipulated. For example, the average power consumption of the motors during the manipulation phase will differ depending on an object’s weight. This data is then clustered using the standard k-means algorithm [MacQueen, 1967] and the resulting clusters are labeled.

Prior to clustering, each element of a data vector is normalized using

$$x_i' = \frac{x_i - \overline{x}_i}{\sigma_i} \tag{1}$$

where $i = 1, \cdots, L$ and L is the length of a data vector $\vec{X}$. The mean $\overline{x}_i$ and variance σ_i^2 are calculated with respect to the training data using

$$\begin{aligned} \overline{x}_i &= \tfrac{1}{N} \textstyle\sum_{n=1}^{N} x_i^n \\ \sigma_i^2 &= \tfrac{1}{N-1} \textstyle\sum_{n=1}^{N} (x_i^n - \overline{x}_i)^2 \end{aligned} \quad . \tag{2}$$

where N is the number of data vectors in the training set. This normalization process is necessary since the elements of a data vector typically have magnitudes that differ significantly.

The labeled clusters are then used to train a radial basis function network [Bishop, 1995] (a subsystem of the `INPUT` module) for classifying the clusters based on proprioceptive and exteroceptive data. Rather than choosing a subset of data points of the clusters as the centers of basis functions, we use the k-means clustering algorithm (in which the number of centers must be decided in advance) to determine for each cluster a set of centers which more accurately reflects the distribution of the cluster’s data points. The appropriate number of center points is determined by the performance of the resulting network on a validation set. In the implemented neural network, we used a Gaussian function as a basis function. Figure 3 shows the topology of the radial basis function network used for data classification.

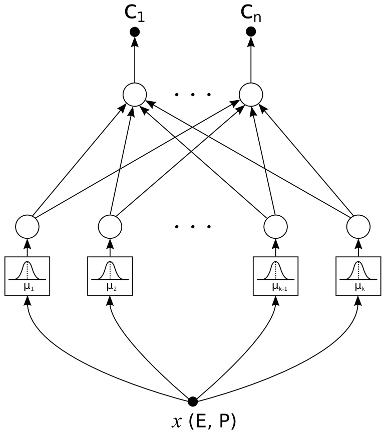

Figure 3: Radial basis function network

4 Experimental Setup

The robot used for testing our system was developed in our group, and is based on the design of the ARAMIES robot [Hilljegerdes *et al.*, 2005]. Our robot is a fully functional ambulating robot that is robust and kinematically flexible. It is equipped with various sensors that enable it to perceive both proprioceptive and exteroceptive signals. On each of the robot’s legs, there are 6 D.C. motors, 6 pressure sensors, and an infrared sensor. For our experiment, the camera of the robot was used as a source of exteroceptive data, and the average motor current consumption of each motor was used as a source of proprioceptive data.

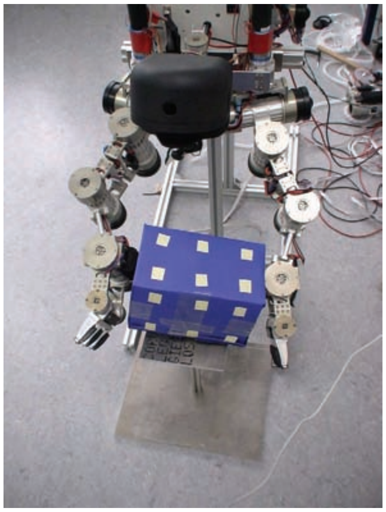

Figure 4: The robot manipulating an object

In the experiment we performed, the robot's body is fixed and it uses its forelegs to manipulate the different objects shown in Figure 5. The objects have differing weights and visual features. Two of the objects have the same visual features, and cannot be distinguished from each other using only visual information; these objects are marked as "A" and "B" in Figure 5. The faces on which the letters are written are placed away from the camera of the robot so that the two objects appear indistinguishable to the robot.

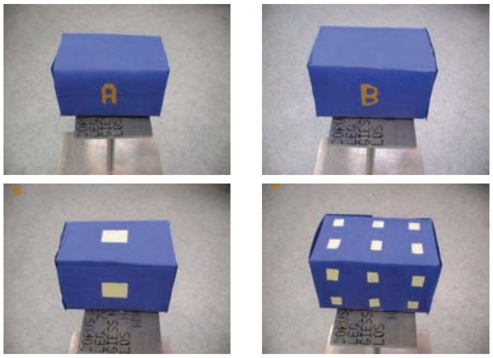

Figure 5: Objects used in the sensorimotor-coordination experiment

In the training session, five manipulation acts were performed on each of the objects. For a single manipulation act, we took a series of images from which we calculated the average number of contours extracted and the average area of the extracted contours. Furthermore, we calculated the total current consumption average for the motors on both of the robot's forelegs.

5 Results

5.1 Repeatability of Features

Table 1 shows, for each of the object, the average and standard deviation of the number of detected contours N_c, the area (number of pixels) of the detected contours A_c, and the total current consumption I (in mA) of both of the robot's forelegs over all training sessions. This data is an indirect measure of the repeatability of a particular feature's measurements. A measurement for a feature is repeatable if the variance of the measurement over a given sample of measurements is small enough that the overlap of measurements resulting from different objects is minimal. One can easily see that the number of contours detected is the most stable feature in this experiment. For getting the number of contours, we used a detector which is robust against noise and changes in lighting conditions. The average current consumption of both legs shows the highest variance in relation to the other features since the end effectors of the forelegs do not grab the object at the same point for each training session. This causes the object's center of gravity to shift with respect to the end effector, and thus a variation in the average current consumption is observed.

$Obj.$	$\overline{N_c}$	$\overline{A_c}$	$\overline{\sum I}$	σN_c	σA_c	$\sigma(\sum I)$
1	1	4812.2	4688.68	0	48.22	217.93
2	1	4925.77	5242.52	0	61.53	159.14
3	2	3134.15	4670.66	0	39.5	181.27
4	6.96	953.1	4916.75	0.21	10.4	319.41

Table 1: The average and standard deviation of features over the whole training set

5.2 Recognition Rates

We tested the system's ability to recognize the objects it was trained for. The system was tested in three different scenarios. In the first scenario, the system was permitted to use both exteroceptive and proprioceptive data to recognize objects. In this case, the recognition rate was the highest, yielding only one misclassification in 20 trials. In the case where the system was allowed to use only exteroceptive data, there were 7 misclassifications in 20 trials. This poorer performance is explained by the fact that two of the objects have the same visual features. In contrast to these results, when only proprioceptive data was used, there were only 3 misclassifications in 20 trials because the weights of each object were unique. An interesting point is that the system was able to correctly classify objects "A" and "B" in this case, which would have caused problems when using only exteroceptive data.

6 Conclusion and Outlook

An embodied recognition system has been presented which learns to recognize objects by interacting with them. We have shown that a learning system trained based on multimodal sensory information can recognize objects by using only partially available (i.e. only exteroceptive, or only proprioceptive) sensory data. The direct byproduct of such systems is a robust system which continues to operate in the absence of either the proprioceptive or the exteroceptive data. Our preliminary results demonstrate that this method can be effectively used in a robotic system which learns from experience about its environment.

In the future, we plan to extend the system by increasing the number of proprioceptive and exteroceptive object-features extracted from the environment, and improving their stability. For example, we may use local features such as SIFT (Scale Invariant Feature Transform) features [Love, 2004] that describe objects distinctively and which are stable against translation, rotation, scaling and different lighting conditions. Moreover, we want to extend the ACTIVATION module of Figure 1 using adaptive manipulation techniques that result in better manipulation skills [Kamon *et al.*, 1998; Coelho *et al.*, 2001; Morales *et al.*, 2004].

References

[Beer, 1996] R. D. Beer. Towards the evolution of a dynamical neural networks for minimally cognitive behavior. In *Proceeding of the 4th International Conference on Simulation of Adaptive Behavior*, pages 421–429, 1996.

[Bishop, 1995] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[Brooks, 1986] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, 1986.

[Coelho *et al.*, 2001] J. Coelho, J. Piater, and R. Grupen. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robots. *Robotics and Autonomous Systems. Special Issue on Humanoid Robots*, 37(2-3):195–218, 2001.

[Edelman, 1987] G. E. Edelman. *Neural Darwinisim: The Theory of neuronal Group Selection.* Basic Books, New York, 1987.

[Fodor, 1983] J. A. Fodor. *The Modularity of Mind.* MIT Press, Massachusetts, London, 1983.

[Gibson, 1977] J.J. Gibson. The theory of affordances. In R. Shaw and J. Brandsford, editors, *Perceiving, Acting, and Knowing: Toward and Ecological Psychology*, pages 62–82. Hillsdalle, 1977.

[Hilljegerdes *et al.*, 2005] J. Hilljegerdes, D. Spenneberg, and F. Kirchner. The construction of the four legged prototype robot aramies. In *Proceedings of the 8th International Conference on Climbing and Walking Robots (CLAWAR 2005)*, London, UK, September 2005.

[James, 1890] W. James. *The Principles of Psychology*. Henry Holt, New York, 1890.

[Kamon *et al.*, 1998] I. Kamon, T. Flash, and S. Edelman. Learning visually guided grasping: a test case in sensorimotor learning. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 28:266–276, 1998.

[Khatib, 1985] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *IEEE International Conference on Robotics and Automation*, pages 505–505, 1985.

[Kirchner *et al.*, 2002] F. Kirchner, D. Spenneberg, and R. Linnemann. A biologically inspired approach towards robust real world locomotion in an 8-legged robot. In J. Ayers, J. Davis, and A. Rudolph, editors, *Neurotechnology for Biomimetic Robots*. MIT-Press, Cambridge, MA, USA, 2002.

[Lakoff and Johnson, 1980] G. Lakoff and M. Johnson. *Metaphors We Live By*. University of Chicago Press, Chicago, 1980.

[Love, 2004] D. Love. Distinctive image features from scale invariant features. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[Lungarella *et al.*, 2005] M. Lungarella, T. Pegors, D. Bulwinkle, and O. Sporns. Methods for quantifying the information structure of sensory and motor data. *Neuroinformatics*, 3(3):243–262, 2005.

[MacQueen, 1967] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, page 281297, 1967.

[Morales *et al.*, 2004] A. Morales, E. Chinellato, A. H. Fagg, and A. P. del Pobil. Active learning for robot manipulation. In *European Conference on Artificial Intelligence (ECAI 2004)*, Valencia, Spain, 2004.

[Nilsson, 1984] N. J. Nilsson. Shakey the robot. SRI Technical Note 323, Menlo Park, CA, USA, 1984.

[Nolfi and Parisi, 1999] S. Nolfi and D. Parisi. Exploiting the power of sensory-motor coordination. In F. Mondada D. Floreano, J-D. Nicoud, editor, *Advances in Artificial Life, Proceedings of Firth European Conference on Artificial Life (ECAL)*, pages 173–182. Springer Verlag, 1999.

[Nolfi, 1996] S. Nolfi. Adaptation as a more powerful tool than decomposition and integration. In *Proceeding of the Workshop on Evolutionary Computing and Machine Learning*, 1996.

[Pfeifer and Scheier, 1999] R. Pfeifer and C. Scheier. *Understanding Intelligence.* MIT Press, Massachusetts, London, 1999.

[Prinz, 1987] W. Prinz. Perspectives on perception and action. In H. Heuer and A. F. Sanders, editors, *Ideo-motor action*, pages 47–76. Lawrence Erlbaum Associates, Hillsdalle, 1987.

[Spenneberg and Kirchner, 2005] D. Spenneberg and F. Kirchner. Learning spatial categories on the basis of proprioceptive data. In *Proceedings of the 3rd International Symposium on Adaptive Motion in Animals and Machines*, September 2005.

[Takác, 2006] M. Takác. Categorization by sensory-motor interaction in artificial agents. In *Proceedings of the 7th International Conference on Cognitive Modeling*, 2006.

[Wilson, 2002] M. Wilson. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636, April 2002.